GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 22:21:32 ; Search time 1180 Seconds
(without alignments)
320.624 Million cell updates/sec

Title: US-09-897-776A-18
Perfect score: 13
Sequence: 1 atggcatggcatg 13

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 804208

Minimum DB seq length: 13
Maximum DB seq length: 50

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		11.4	87.7	14	6	AR131417	AR131417 Sequence
2	c	11.4	87.7	14	6	AR131417	AR131417 Sequence
3		11.4	87.7	14	6	AR148643	AR148643 Sequence
4	c	11.4	87.7	14	6	AR148643	AR148643 Sequence
5		11.4	87.7	14	6	AR154223	AR154223 Sequence
6	c	11.4	87.7	14	6	AR154223	AR154223 Sequence
7		11.4	87.7	14	6	AX259319	AX259319 Sequence
8	c	11.4	87.7	14	6	AX259319	AX259319 Sequence
9		11.4	87.7	14	6	AX339614	AX339614 Sequence
10	c	11.4	87.7	14	6	AX339614	AX339614 Sequence
11		11.4	87.7	14	6	I40291	I40291 Sequence 1
12	c	11.4	87.7	14	6	I40291	I40291 Sequence 1
13		11.4	87.7	14	6	I76128	I76128 Sequence 1
14	c	11.4	87.7	14	6	I76128	I76128 Sequence 1
15	c	11.4	87.7	20	6	AR029339	AR029339 Sequence
16		11.4	87.7	20	6	AR123187	AR123187 Sequence
17		11.4	87.7	20	6	AR157706	AR157706 Sequence
18		11.4	87.7	24	6	AX447375	AX447375 Sequence
19	c	11.4	87.7	30	6	AR083983	AR083983 Sequence
20	c	11.4	87.7	30	6	AR160899	AR160899 Sequence
21		11.4	87.7	31	6	AR094401	AR094401 Sequence
22	c	11.4	87.7	31	6	AR094401	AR094401 Sequence
23		11.4	87.7	31	6	AR098615	AR098615 Sequence
24	c	11.4	87.7	31	6	AR098615	AR098615 Sequence
25		11.4	87.7	32	6	AX076709	AX076709 Sequence
26		11.4	87.7	34	6	A75796	A75796 Sequence 12
27		11.4	87.7	36	6	AX356328	AX356328 Sequence
28	c	11	84.6	31	6	AX248632	AX248632 Sequence
29		10.4	80.0	14	6	A88232	A88232 Sequence 38
30		10.4	80.0	14	6	A90199	A90199 Sequence 38
31		10.4	80.0	18	6	AR036701	AR036701 Sequence
32	c	10.4	80.0	18	6	AR211219	AR211219 Sequence
33	c	10.4	80.0	18	6	AX060758	AX060758 Sequence
34	c	10.4	80.0	18	6	AX060937	AX060937 Sequence
35	c	10.4	80.0	20	6	A22793	A22793 E1B HCV pro
36	c	10.4	80.0	20	6	AR031220	AR031220 Sequence
37		10.4	80.0	20	6	AR099484	AR099484 Sequence
38	c	10.4	80.0	20	6	AR145036	AR145036 Sequence
39		10.4	80.0	20	6	AR178765	AR178765 Sequence
40		10.4	80.0	20	6	AR178933	AR178933 Sequence
41		10.4	80.0	20	6	E04295	E04295 DNA encodin
42	c	10.4	80.0	21	6	I73432	I73432 Sequence 37
43	c	10.4	80.0	28	6	AX203607	AX203607 Sequence
44		10.4	80.0	30	6	AR160899	AR160899 Sequence
45	c	10.4	80.0	33	6	AR004364	AR004364 Sequence

ALIGNMENTS

RESULT 1
AR131417
LOCUS AR131417 14 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6194144.
ACCESSION AR131417
VERSION AR131417.1 GI:14120320
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 14)
AUTHORS Koster,H.
TITLE DNA sequencing by mass spectrometry
JOURNAL Patent: US 6194144-A 1 27-FEB-2001;
FEATURES Location/Qualifiers

```
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db      2 ATGCCATGGCATG 14


RESULT 2
AR131417/c
LOCUS       AR131417                  14 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 1 from patent US 6194144.
ACCESSION   AR131417
VERSION     AR131417.1  GI:14120320
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Koster,H.
  TITLE     DNA sequencing by mass spectrometry
  JOURNAL   Patent: US 6194144-A 1 27-FEB-2001;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db     13 ATGCCATGGCATG 1


RESULT 3
AR148643
LOCUS       AR148643                  14 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 1 from patent US 6225450.
ACCESSION   AR148643
VERSION     AR148643.1  GI:15112733
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Koster,H.
  TITLE     DNA sequencing by mass spectrometry
  JOURNAL   Patent: US 6225450-A 1 01-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db      2 ATGCCATGGCATG 14


RESULT 4
```

```
AR148643/c
LOCUS       AR148643                  14 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 1 from patent US 6225450.
ACCESSION   AR148643
VERSION     AR148643.1  GI:15112733
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Koster,H.
  TITLE     DNA sequencing by mass spectrometry
  JOURNAL   Patent: US 6225450-A 1 01-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db     13 ATGCCATGGCATG 1


RESULT 5
AR154223
LOCUS       AR154223                  14 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 1 from patent US 6238871.
ACCESSION   AR154223
VERSION     AR154223.1  GI:15122276
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Koster,H.
  TITLE     DNA sequences by mass spectrometry
  JOURNAL   Patent: US 6238871-A 1 29-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db      2 ATGCCATGGCATG 14


RESULT 6
AR154223/c
LOCUS       AR154223                  14 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 1 from patent US 6238871.
ACCESSION   AR154223
VERSION     AR154223.1  GI:15122276
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Koster,H.
  TITLE     DNA sequences by mass spectrometry
  JOURNAL   Patent: US 6238871-A 1 29-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .14
```

```
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db     13 ATGCCATGGCATG 1


RESULT 7
AX259319
LOCUS       AX259319                  14 bp    DNA     linear   PAT 26-OCT-2001
DEFINITION  Sequence 1 from Patent WO0173085.
ACCESSION   AX259319
VERSION     AX259319.1  GI:16508556
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Duering,K.
  TITLE     Expression vectors for concentrating a recombinantly produced
            protein in different cell compartments
  JOURNAL   Patent: WO 0173085-A 1 04-OCT-2001;
            MPB Cologne GmbH Molecular Plant & Protein Biotechnology (DE)
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligonukleotid"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db      2 ATGCCATGGCATG 14


RESULT 8
AX259319/c
LOCUS       AX259319                  14 bp    DNA     linear   PAT 26-OCT-2001
DEFINITION  Sequence 1 from Patent WO0173085.
ACCESSION   AX259319
VERSION     AX259319.1  GI:16508556
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Duering,K.
  TITLE     Expression vectors for concentrating a recombinantly produced
            protein in different cell compartments
  JOURNAL   Patent: WO 0173085-A 1 04-OCT-2001;
            MPB Cologne GmbH Molecular Plant & Protein Biotechnology (DE)
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligonukleotid"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db     13 ATGCCATGGCATG 1


RESULT 9
AX339614
LOCUS       AX339614                  14 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 6 from Patent WO0196551.
ACCESSION   AX339614
VERSION     AX339614.1  GI:18135627
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Short,J.M.
  TITLE     Whole cell engineering by mutagenizing a substantial portion of a
            starting genome, combining mutations, and optionally repeating
  JOURNAL   Patent: WO 0196551-A 6 20-DEC-2001;
            DIVERSA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Tetradecanucleotide d"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db      2 ATGCCATGGCATG 14


RESULT 10
AX339614/c
LOCUS       AX339614                  14 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 6 from Patent WO0196551.
ACCESSION   AX339614
VERSION     AX339614.1  GI:18135627
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Short,J.M.
  TITLE     Whole cell engineering by mutagenizing a substantial portion of a
            starting genome, combining mutations, and optionally repeating
  JOURNAL   Patent: WO 0196551-A 6 20-DEC-2001;
            DIVERSA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Tetradecanucleotide d"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCATG 13
          ||| |||||||||
Db     13 ATGCCATGGCATG 1


RESULT 11
```

```
I40291
LOCUS       I40291                    14 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 1 from patent US 5620849.
ACCESSION   I40291
VERSION     I40291.1  GI:2082583
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Botchan,M.R., Yang,L., Li,R., Mohr,I.J. and Clark,R.
  TITLE     Methods and compositions for identifying inhibitors of papilloma
            virus replication
  JOURNAL   Patent: US 5620849-A 1 15-APR-1997;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGGCATGGCATG 13
           ||| |||||||||
Db       2 ATGCCATGGCATG 14


RESULT 12
I40291/c
LOCUS       I40291                    14 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 1 from patent US 5620849.
ACCESSION   I40291
VERSION     I40291.1  GI:2082583
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Botchan,M.R., Yang,L., Li,R., Mohr,I.J. and Clark,R.
  TITLE     Methods and compositions for identifying inhibitors of papilloma
            virus replication
  JOURNAL   Patent: US 5620849-A 1 15-APR-1997;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGGCATGGCATG 13
           ||| |||||||||
Db      13 ATGCCATGGCATG 1


RESULT 13
I76128
LOCUS       I76128                    14 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 1 from patent US 5691141.
ACCESSION   I76128
VERSION     I76128.1  GI:3012282
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Koster,H.
  TITLE     DNA sequencing by mass spectrometry
  JOURNAL   Patent: US 5691141-A 1 25-NOV-1997;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGGCATGGCATG 13
           ||| |||||||||
Db       2 ATGCCATGGCATG 14


RESULT 14
I76128/c
LOCUS       I76128                    14 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 1 from patent US 5691141.
ACCESSION   I76128
VERSION     I76128.1  GI:3012282
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Koster,H.
  TITLE     DNA sequencing by mass spectrometry
  JOURNAL   Patent: US 5691141-A 1 25-NOV-1997;
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="unknown"
BASE COUNT        3 a      4 c      4 g      3 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGGCATGGCATG 13
           ||| |||||||||
Db      13 ATGCCATGGCATG 1


RESULT 15
AR029339/c
LOCUS       AR029339                  20 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 61 from patent US 5859230.
ACCESSION   AR029339
VERSION     AR029339.1  GI:5941312
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Kim,J.P., Reyes,G.R. and Young,L.Marie.
  TITLE     Non-A/non-B/non-C/non-D/non-E hepatitis agents and molecular
            cloning thereof
  JOURNAL   Patent: US 5859230-A 61 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        5 a      7 c      4 g      4 t
ORIGIN

  Query Match             87.7%;  Score 11.4;  DB 6;  Length 20;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+04;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGGCATGGCATG 13
           |||||||||| ||
Db      14 ATGGCATGGCTTG 2
```

Search completed: March 30, 2003, 00:16:56
Job time : 1182 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 13:49:51 ; Search time 3576 Seconds
(without alignments)
8.187 Million cell updates/sec

Title: US-09-897-776A-18
Perfect score: 13
Sequence: 1 atggcatggcatg 13

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 1838882

Minimum DB seq length: 13
Maximum DB seq length: 50

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	12	92.3	35	22	AAH20938	pRT100/FRT-scFv(ox
c	2	12	92.3	35	24	AAI69338	FLP-recombinase-LB
	3	11.4	87.7	14	22	AAH49474	scFv(ox) antibody
c	4	11.4	87.7	14	22	AAH49474	scFv(ox) antibody
	5	11.4	87.7	14	22	AAH49492	Endoplasmic reticu
c	6	11.4	87.7	14	22	AAH49492	Endoplasmic reticu
	7	11.4	87.7	14	22	AAH44110	Tetradecanucleotid
c	8	11.4	87.7	14	22	AAH44110	Tetradecanucleotid
	9	11.4	87.7	14	22	AAH20936	Anaerobically-indu
c	10	11.4	87.7	14	22	AAH20936	Anaerobically-indu
	11	11.4	87.7	14	24	AAI69334	Plasmid pRT100/scF
c	12	11.4	87.7	14	24	AAI69334	Plasmid pRT100/scF
c	13	11.4	87.7	18	21	AAZ70542	Human biallelic ma
	14	11.4	87.7	20	17	AAT39773	Target sequence fo
c	15	11.4	87.7	20	17	AAT07059	Primer 17A-215F fo
	16	11.4	87.7	21	24	ABL44152	Human chromosome 1
	17	11.4	87.7	24	24	ABQ03823	Oligonucleotide ad
c	18	11.4	87.7	30	19	AAV69346	Corn glycogenin PC
c	19	11.4	87.7	30	20	AAV82034	Moraxella lactofer
c	20	11.4	87.7	30	22	AAD09246	Corn glycogenin cc
c	21	11.4	87.7	30	24	ABK15087	Corn glycogenin-sp
	22	11.4	87.7	31	18	AAT58149	PCR primer used to
c	23	11.4	87.7	31	18	AAT58149	PCR primer used to
	24	11.4	87.7	32	22	AAF56168	Streptomyces caele
	25	11.4	87.7	34	14	AAQ50943	PCR primer for amp
c	26	11.4	87.7	34	20	AAZ29952	PCR primer DYV-099
	27	11.4	87.7	36	24	ABA05920	Escherichia coli y
	28	11.4	87.7	50	22	AAL33858	Human SNP oligonuc
c	29	11	84.6	22	19	AAV30000	Human cell growth
	30	11	84.6	50	22	AAL32466	Human SNP oligonuc
	31	10.6	81.5	15	24	ABL91819	Human LIPG gene al
	32	10.4	80.0	14	19	AAV48791	ErbB-2 gene antise
	33	10.4	80.0	15	22	AAF52444	IGF-I oligonucleot
	34	10.4	80.0	15	22	AAF52445	IGF-I oligonucleot
	35	10.4	80.0	15	22	AAF52446	IGF-I oligonucleot
	36	10.4	80.0	15	22	AAF52447	IGF-I oligonucleot
	37	10.4	80.0	17	11	AAQ06575	Probe 1 to detect
c	38	10.4	80.0	18	21	AAA55621	TRAF4 antisense ol
c	39	10.4	80.0	18	24	AAS98449	Human protective D
c	40	10.4	80.0	19	21	AAZ70056	Human biallelic ma
	41	10.4	80.0	19	24	ABA02044	Serine pathway met
	42	10.4	80.0	20	13	AAQ29663	PCR primer #129 fo
	43	10.4	80.0	20	13	AAQ31492	NANB hepatitis vir
c	44	10.4	80.0	20	13	AAQ32975	HCV primer E1B (-)
	45	10.4	80.0	20	14	AAQ43875	NANB hepatitis vir

ALIGNMENTS

RESULT 1
AAH20938/c
ID AAH20938 standard; DNA; 35 BP.
XX
AC AAH20938;
XX
DT 24-AUG-2001 (first entry)
XX
DE pRT100/FRT-scFv(ox) primer NcoI-PLP-LBD.
XX
KW Primer; anaerobic; promoter; GapC4; transgenic; inducer; ss.
XX
OS Unidentified.
XX
PN WO200138508-A2.
XX
PD 31-MAY-2001.
XX
PF 05-SEP-2000; 2000WO-DE03119.
XX
PR 23-NOV-1999; 99DE-1056272.
XX
PA (MPBC-) MPB COLOGNE GMBH MOLECULAR PLANT & PROTE.
XX
PI Duering K, Buelow L;
XX
DR WPI; 2001-367680/38.
XX
PT Producing proteins from transgenic organisms after harvesting, useful
PT e.g. for preparing single-chain antibodies, by chemical induction of
PT protein-expressing gene -
XX

PS Example 2; Page 17; 22pp; German.
XX
CC This invention describes a novel method for the production of a selected
CC protein (I) by a transgenic organism (A) in which expression of the
CC (I)-encoding gene (II) occurs only after harvesting of (A). (A) contains
CC (II) that is expressed only in the presence of a chemical inducer (III)
CC and harvested (A) are contacted with (III) by delivering this to a phase
CC that surrounds (A). The method is used to produce (I), or other
CC (bio)chemicals, on a large scale for (bio)medical (therapeutic or
CC diagnostic) purposes or industrial applications. The examples illustrate
CC production of single-chain Fv antibody fragments. Expression of (I) only
CC after harvest eliminates the need to comminute tissue and (III) can be
CC delivered simply and uniformly to the cells of (A). This sequence
CC represents a primer derived from an anaerobically induced GapC4 promoter
CC which is used to illustrate the method of the invention.
XX
SQ Sequence 35 BP; 13 A; 7 C; 4 G; 11 T; 0 other;

Query Match 92.3%; Score 12; DB 22; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGGCATGGCATG 13
||||||||||||
Db 12 TGGCATGGCATG 1

RESULT 2
AAI69338/c
ID AAI69338 standard; DNA; 35 BP.
XX
AC AAI69338;
XX
DT 18-FEB-2002 (first entry)
XX
DE FLP-recombinase-LBD fusion construct primer HincII-pGapC4.
XX
KW Protein farming; transgenic plant; antimicrobial; antibiotic; RNAase;
KW diphtheria toxin; cytosine deaminase; polyhydroxyalkanoate production;
KW fusion construct; primer; ss.
XX
OS Synthetic.
XX
PN WO200188164-A1.
XX
PD 22-NOV-2001.
XX
PF 28-FEB-2001; 2001WO-DE00780.
XX
PR 19-MAY-2000; 2000DE-1024740.
XX
PA (MPBC-) MPB COLOGNE GMBH.
XX
PI Duering K;
XX
DR WPI; 2002-055703/07.
XX
PT Controlled elimination of DNA, useful for expressing toxic proteins in
PT plants, comprises expressing recombinase-ligand binding domain fusion
PT from an inducible promoter -
XX
PS Example 1; Page 19; 31pp; German.
XX
CC This invention describes a novel method for the controlled elimination of
CC a selected DNA sequence (I) from a host organism. The host is transformed
CC with the following: (i) a DNA sequence (Ia), flanked by 5' and 3'
CC recombination DNA sequences (RS); and (ii) a sequence (II) that encodes a
CC ligand-binding domain/recombinase fusion protein (Rec-LBD) that
CC recognizes RS, under control of an inducible promoter (iP).
CC Transformation is under conditions where iP is repressed (no Rec-LBD is
CC produced), after which iP is induced to cause expression of Rec-LBD and
CC a ligand is added to activate this fusion protein. The method is
CC applicable to plants to provide temporally regulated expression of
CC foreign proteins ('protein farming'), particularly for post-harvest
CC recovery of the proteins. The proteins are particularly toxic or
CC deleterious to the plant, e.g. antimicrobial or antibiotic peptides,
CC RNAases, diphtheria toxin, cytosine deaminase, antibodies etc. The method
CC may also be used to remove marker genes from transgenic plants and for
CC production of poly(hydroxyalkanoates). The method includes two levels of
CC repression (use of inducible promoter and ligand for activating the
CC recombination reaction), ensuring secure regulation of recombinase
CC activity as long as this is required (generally until plants are
CC harvested). By selection of appropriate promoters, tissue selectivity may
CC also be provided. This sequence represents a primer used in the
CC construction of FLP-recombinase-ligand binding domain (LBD) fusion
CC protein encoding DNA.
XX
SQ Sequence 35 BP; 13 A; 7 C; 4 G; 11 T; 0 other;

Query Match 92.3%; Score 12; DB 24; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGGCATGGCATG 13
||||||||||||
Db 12 TGGCATGGCATG 1

RESULT 3
AAH49474
ID AAH49474 standard; DNA; 14 BP.
XX
AC AAH49474;
XX
DT 11-DEC-2001 (first entry)
XX
DE scFv(ox) antibody KDEL-ER targetting sequence linker DNA fragment.
XX
KW Antibody scFV(oX); fusion protein; localization signal; plant;
KW potato; ss.
XX
OS Solanum tuberosum.
OS Synthetic.
XX
PN DE10014412-A1.
XX
PD 04-OCT-2001.
XX
PF 24-MAR-2000; 2000DE-1014412.
XX
PR 24-MAR-2000; 2000DE-1014412.
XX
PA (MPBC-) MPB COLOGNE GMBH MOLECULAR PLANT & PROTE.
XX
PI Duering K;
XX
DR WPI; 2001-607899/70.
XX
PT Expression vector comprising several copies of the same gene, useful
PT for expressing proteins in plants, in several different cell
PT compartments -
XX
PS Example 1; Column 9; 10pp; German.
XX
CC This invention describes a novel expression vector (A) containing at
CC least two copies of a gene (I) encoding a protein (II), each linked to a
CC promoter, or a composition (B) of at least two (A), each with at least
CC one copy of (I) linked to a promoter. The individual (I) encode (II) as
CC a fusion protein with a localization signal (LS), with all (II)-encoding
CC parts being the same but the LS different. After introduction of (A) or
CC (B) into a host, (I) is expressed in different compartments. (A) and (B),
CC or combinations of them, are used for production of (II) in plants
CC (molecular farming). Expressing (I) in several different compartments of
CC plant cells provides increased production of (II). This sequence
CC represents a linker fragment used in the construction of the scFv(ox)
CC antibody described in the method of the invention.

XX
SQ Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

Query Match 87.7%; Score 11.4; DB 22; Length 14;
Best Local Similarity 92.3%; Pred. No. 2.6e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||| |||||||||
Db 2 ATGCCATGGCATG 14

RESULT 4
AAH49474/c
ID AAH49474 standard; DNA; 14 BP.
XX
AC AAH49474;
XX
DT 11-DEC-2001 (first entry)
XX
DE scFv(ox) antibody KDEL-ER targetting sequence linker DNA fragment.
XX
KW Antibody scFV(oX); fusion protein; localization signal; plant;
KW potato; ss.
XX
OS Solanum tuberosum.
OS Synthetic.
XX
PN DE10014412-A1.
XX
PD 04-OCT-2001.
XX
PF 24-MAR-2000; 2000DE-1014412.
XX
PR 24-MAR-2000; 2000DE-1014412.
XX
PA (MPBC-) MPB COLOGNE GMBH MOLECULAR PLANT & PROTE.
XX
PI Duering K;
XX
DR WPI; 2001-607899/70.
XX
PT Expression vector comprising several copies of the same gene, useful
PT for expressing proteins in plants, in several different cell
PT compartments -
XX
PS Example 1; Column 9; 10pp; German.
XX
CC This invention describes a novel expression vector (A) containing at
CC least two copies of a gene (I) encoding a protein (II), each linked to a
CC promoter, or a composition (B) of at least two (A), each with at least
CC one copy of (I) linked to a promoter. The individual (I) encode (II) as
CC a fusion protein with a localization signal (LS), with all (II)-encoding
CC parts being the same but the LS different. After introduction of (A) or
CC (B) into a host, (I) is expressed in different compartments. (A) and (B),
CC or combinations of them, are used for production of (II) in plants
CC (molecular farming). Expressing (I) in several different compartments of
CC plant cells provides increased production of (II). This sequence
CC represents a linker fragment used in the construction of the scFv(ox)
CC antibody described in the method of the invention.
XX
SQ Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

Query Match 87.7%; Score 11.4; DB 22; Length 14;
Best Local Similarity 92.3%; Pred. No. 2.6e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||| |||||||||
Db 13 ATGCCATGGCATG 1

RESULT 5
AAH49492
ID AAH49492 standard; DNA; 14 BP.
XX
AC AAH49492;
XX
DT 11-DEC-2001 (first entry)
XX
DE Endoplasmic reticulum targetting scFv(ox) antibody linker #1.
XX
KW Antibody scFV(oX); fusion protein; localization signal; plant;
KW endoplasmic reticulum; ss.
XX
OS Unidentified.
XX
PN DE10014412-A1.
XX
PD 04-OCT-2001.
XX
PF 24-MAR-2000; 2000DE-1014412.
XX
PR 24-MAR-2000; 2000DE-1014412.
XX
PA (MPBC-) MPB COLOGNE GMBH MOLECULAR PLANT & PROTE.
XX
PI Duering K;
XX
DR WPI; 2001-607899/70.
XX
PT Expression vector comprising several copies of the same gene, useful
PT for expressing proteins in plants, in several different cell
PT compartments -
XX
PS Example 3; Column 13; 10pp; German.
XX
CC This invention describes a novel expression vector (A) containing at
CC least two copies of a gene (I) encoding a protein (II), each linked to a
CC promoter, or a composition (B) of at least two (A), each with at least
CC one copy of (I) linked to a promoter. The individual (I) encode (II) as
CC a fusion protein with a localization signal (LS), with all (II)-encoding
CC parts being the same but the LS different. After introduction of (A) or
CC (B) into a host, (I) is expressed in different compartments. (A) and (B),
CC or combinations of them, are used for production of (II) in plants
CC (molecular farming). Expressing (I) in several different compartments of
CC plant cells provides increased production of (II). This sequence
CC represents an endoplasmic reticulum (ER) scFv(ox) antibody linker
CC fragment described in the method of the invention.
XX
SQ Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

Query Match 87.7%; Score 11.4; DB 22; Length 14;
Best Local Similarity 92.3%; Pred. No. 2.6e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||| |||||||||
Db 2 ATGCCATGGCATG 14

RESULT 6
AAH49492/c
ID AAH49492 standard; DNA; 14 BP.
XX
AC AAH49492;
XX
DT 11-DEC-2001 (first entry)
XX
DE Endoplasmic reticulum targetting scFv(ox) antibody linker #1.
XX
KW Antibody scFV(oX); fusion protein; localization signal; plant;
KW endoplasmic reticulum; ss.
XX
OS Unidentified.
XX

PN DE10014412-A1.
XX
PD 04-OCT-2001.
XX
PF 24-MAR-2000; 2000DE-1014412.
XX
PR 24-MAR-2000; 2000DE-1014412.
XX
PA (MPBC-) MPB COLOGNE GMBH MOLECULAR PLANT & PROTE.
XX
PI Duering K;
XX
DR WPI; 2001-607899/70.
XX
PT Expression vector comprising several copies of the same gene, useful
PT for expressing proteins in plants, in several different cell
PT compartments -
XX
PS Example 3; Column 13; 10pp; German.
XX
CC This invention describes a novel expression vector (A) containing at
CC least two copies of a gene (I) encoding a protein (II), each linked to a
CC promoter, or a composition (B) of at least two (A), each with at least
CC one copy of (I) linked to a promoter. The individual (I) encode (II) as
CC a fusion protein with a localization signal (LS), with all (II)-encoding
CC parts being the same but the LS different. After introduction of (A) or
CC (B) into a host, (I) is expressed in different compartments. (A) and (B),
CC or combinations of them, are used for production of (II) in plants
CC (molecular farming). Expressing (I) in several different compartments of
CC plant cells provides increased production of (II). This sequence
CC represents an endoplasmic reticulum (ER) scFv(ox) antibody linker
CC fragment described in the method of the invention.
XX
SQ Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

Query Match 87.7%; Score 11.4; DB 22; Length 14;
Best Local Similarity 92.3%; Pred. No. 2.6e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 ATGGCATGGCATG 13
           ||| |||||||||
Db      13 ATGCCATGGCATG 1
```

RESULT 7
AAH44110
ID AAH44110 standard; DNA; 14 BP.
XX
AC AAH44110;
XX
DT 13-SEP-2001 (first entry)
XX
DE Tetradecanucleotide SEQ ID NO:1.
XX
KW DNA sequencing; Sanger sequencing; base-specific chain termination;
KW mass spectrometry; hybridisation; probe; tag; molecular weight; ss.
XX
OS Synthetic.
XX
PN US6225450-B1.
XX
PD 01-MAY-2001.
XX
PF 07-JUN-1995; 95US-0481033.
XX
PR 06-JAN-1994; 94US-0178216.
PR 07-JAN-1993; 93US-0001323.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Koester H;
XX
DR WPI; 2001-482100/52.
XX
PT Mass-modified nucleic acid molecules for DNA and RNA Sanger sequencing,
PT have positively mass-modified nucleotides containing halogens or
PT functional groups attached to heterocyclic base or sugar moiety of
PT nucleotide -
XX
PS Disclosure; Column 6; 61pp; English.
XX
CC The present invention describes an intact ionised and volatilised
CC mass-modified (MM) nucleic acid molecule (I), comprising at least two MM
CC nucleotides chosen from MM 2'-deoxynucleotide, MM 2' 3'-
CC dideoxynucleotide, MM nucleotide and a MM 3'-deoxynucleotide, which are
CC different from each other and positively charged. (I) comprises a MM
CC universal primer or a MM initiator oligonucleotide. Also described are:
CC (1) a set of mass-differentiated tag probes, where each tag probe in
CC the set comprises a sequence of nucleotides which is complementary by
CC Watson-Crick base pairing to a tag sequence present within a set of
CC base-specifically terminated fragments; and (2) an ionised positively
CC charged intact duplex, comprising a MM tag probe having a MM nucleotide
CC bound to a tag sequence present with a base-specifically terminated
CC nucleic acid fragment. The MM nucleotides are useful for DNA and RNA
CC Sanger sequencing. Introducing mass modifications in the oligonucleotide
CC primer, chain-terminating nucleoside triphosphates and/or in the
CC chain-elongating nucleoside triphosphates allows multiplexing by
CC hybridisation tag specific probes with mass differentiated molecular
CC weights. The MM oligonucleotide provides a new method to sequence DNA
CC over the existing DNA sequencing techniques including high speed, high
CC throughput, no requirement of electrophoresis and no costly reagents
CC involving various substitutions with stable isotopes. The present
CC sequence represents a tetradecanucleotide which is given in the
CC exemplification of the present invention.
XX
SQ Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

Query Match 87.7%; Score 11.4; DB 22; Length 14;
Best Local Similarity 92.3%; Pred. No. 2.6e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 ATGGCATGGCATG 13
           ||| |||||||||
Db       2 ATGCCATGGCATG 14
```

RESULT 8
AAH44110/c
ID AAH44110 standard; DNA; 14 BP.
XX
AC AAH44110;
XX
DT 13-SEP-2001 (first entry)
XX
DE Tetradecanucleotide SEQ ID NO:1.
XX
KW DNA sequencing; Sanger sequencing; base-specific chain termination;
KW mass spectrometry; hybridisation; probe; tag; molecular weight; ss.
XX
OS Synthetic.
XX
PN US6225450-B1.
XX
PD 01-MAY-2001.
XX
PF 07-JUN-1995; 95US-0481033.
XX
PR 06-JAN-1994; 94US-0178216.
PR 07-JAN-1993; 93US-0001323.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Koester H;
XX
DR WPI; 2001-482100/52.
XX

PT Mass-modified nucleic acid molecules for DNA and RNA Sanger sequencing,
PT have positively mass-modified nucleotides containing halogens or
PT functional groups attached to heterocyclic base or sugar moiety of
PT nucleotide -
XX
PS Disclosure; Column 6; 61pp; English.
XX
CC The present invention describes an intact ionised and volatilised
CC mass-modified (MM) nucleic acid molecule (I), comprising at least two MM
CC nucleotides chosen from MM 2'-deoxynucleotide, MM 2' 3'-
CC dideoxynucleotide, MM nucleotide and a MM 3'-deoxynucleotide, which are
CC different from each other and positively charged. (I) comprises a MM
CC universal primer or a MM initiator oligonucleotide. Also described are:
CC (1) a set of mass-differentiated tag probes, where each tag probe in
CC the set comprises a sequence of nucleotides which is complementary by
CC Watson-Crick base pairing to a tag sequence present within a set of
CC base-specifically terminated fragments; and (2) an ionised positively
CC charged intact duplex, comprising a MM tag probe having a MM nucleotide
CC bound to a tag sequence present with a base-specifically terminated
CC nucleic acid fragment. The MM nucleotides are useful for DNA and RNA
CC Sanger sequencing. Introducing mass modifications in the oligonucleotide
CC primer, chain-terminating nucleoside triphosphates and/or in the
CC chain-elongating nucleoside triphosphates allows multiplexing by
CC hybridisation tag specific probes with mass differentiated molecular
CC weights. The MM oligonucleotide provides a new method to sequence DNA
CC over the existing DNA sequencing techniques including high speed, high
CC throughput, no requirement of electrophoresis and no costly reagents
CC involving various substitutions with stable isotopes. The present
CC sequence represents a tetradecanucleotide which is given in the
CC exemplification of the present invention.
XX
SQ Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

Query Match 87.7%; Score 11.4; DB 22; Length 14;
Best Local Similarity 92.3%; Pred. No. 2.6e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||| |||||||||
Db 13 ATGCCATGGCATG 1

RESULT 9
AAH20936
ID AAH20936 standard; DNA; 14 BP.
XX
AC AAH20936;
XX
DT 24-AUG-2001 (first entry)
XX
DE Anaerobically-induced GapC4 promoter associated primer #1.
XX
KW Primer; anaerobic; promoter; GapC4; transgenic; inducer; ss.
XX
OS Unidentified.
XX
PN WO200138508-A2.
XX
PD 31-MAY-2001.
XX
PF 05-SEP-2000; 2000WO-DE03119.
XX
PR 23-NOV-1999; 99DE-1056272.
XX
PA (MPBC-) MPB COLOGNE GMBH MOLECULAR PLANT & PROTE.
XX
PI Duering K, Buelow L;
XX
DR WPI; 2001-367680/38.
XX
PT Producing proteins from transgenic organisms after harvesting, useful
PT e.g. for preparing single-chain antibodies, by chemical induction of
PT protein-expressing gene -
XX
PS Example 1; Page 15; 22pp; German.
XX
CC This invention describes a novel method for the production of a selected
CC protein (I) by a transgenic organism (A) in which expression of the
CC (I)-encoding gene (II) occurs only after harvesting of (A). (A) contains
CC (II) that is expressed only in the presence of a chemical inducer (III)
CC and harvested (A) are contacted with (III) by delivering this to a phase
CC that surrounds (A). The method is used to produce (I), or other
CC (bio)chemicals, on a large scale for (bio)medical (therapeutic or
CC diagnostic) purposes or industrial applications. The examples illustrate
CC production of single-chain Fv antibody fragments. Expression of (I) only
CC after harvest eliminates the need to comminute tissue and (III) can be
CC delivered simply and uniformly to the cells of (A). This sequence
CC represents a primer derived from an anaerobically induced GapC4 promoter
CC which is used to illustrate the method of the invention.
XX
SQ Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

Query Match 87.7%; Score 11.4; DB 22; Length 14;
Best Local Similarity 92.3%; Pred. No. 2.6e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||| |||||||||
Db 2 ATGCCATGGCATG 14

RESULT 10
AAH20936/c
ID AAH20936 standard; DNA; 14 BP.
XX
AC AAH20936;
XX
DT 24-AUG-2001 (first entry)
XX
DE Anaerobically-induced GapC4 promoter associated primer #1.
XX
KW Primer; anaerobic; promoter; GapC4; transgenic; inducer; ss.
XX
OS Unidentified.
XX
PN WO200138508-A2.
XX
PD 31-MAY-2001.
XX
PF 05-SEP-2000; 2000WO-DE03119.
XX
PR 23-NOV-1999; 99DE-1056272.
XX
PA (MPBC-) MPB COLOGNE GMBH MOLECULAR PLANT & PROTE.
XX
PI Duering K, Buelow L;
XX
DR WPI; 2001-367680/38.
XX
PT Producing proteins from transgenic organisms after harvesting, useful
PT e.g. for preparing single-chain antibodies, by chemical induction of
PT protein-expressing gene -
XX
PS Example 1; Page 15; 22pp; German.
XX
CC This invention describes a novel method for the production of a selected
CC protein (I) by a transgenic organism (A) in which expression of the
CC (I)-encoding gene (II) occurs only after harvesting of (A). (A) contains
CC (II) that is expressed only in the presence of a chemical inducer (III)
CC and harvested (A) are contacted with (III) by delivering this to a phase
CC that surrounds (A). The method is used to produce (I), or other
CC (bio)chemicals, on a large scale for (bio)medical (therapeutic or
CC diagnostic) purposes or industrial applications. The examples illustrate
CC production of single-chain Fv antibody fragments. Expression of (I) only
CC after harvest eliminates the need to comminute tissue and (III) can be
CC delivered simply and uniformly to the cells of (A). This sequence

```
CC   represents a primer derived from an anaerobically induced GapC4 promoter
CC   which is used to illustrate the method of the invention.
XX
SQ   Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

  Query Match             87.7%;  Score 11.4;  DB 22;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+03;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGGCATGGCATG 13
           ||| |||||||||
Db      13 ATGCCATGGCATG 1


RESULT 11
AAI69334
ID   AAI69334 standard; DNA; 14 BP.
XX
AC   AAI69334;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Plasmid pRT100/scFv(ox)E ER localising scFv(ox) Ab linker primer #1.
XX
KW   Endoplasmic reticulum; scFv; antibody; protein farming; transgenic plant;
KW   antimicrobial; antibiotic; RNAase; diphtheria toxin; cytosine deaminase;
KW   polyhydroxyalkanoate production; primer; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag=  a
FT                   /mod_base= "OTHER"
FT                   /note= "5'-end phosphorylated"
XX
PN   WO200188164-A1.
XX
PD   22-NOV-2001.
XX
PF   28-FEB-2001; 2001WO-DE00780.
XX
PR   19-MAY-2000; 2000DE-1024740.
XX
PA   (MPBC-) MPB COLOGNE GMBH.
XX
PI   Duering K;
XX
DR   WPI; 2002-055703/07.
XX
PT   Controlled elimination of DNA, useful for expressing toxic proteins in
PT   plants, comprises expressing recombinase-ligand binding domain fusion
PT   from an inducible promoter  -
XX
PS   Example 1; Page 18; 31pp; German.
XX
CC   This invention describes a novel method for the controlled elimination of
CC   a selected DNA sequence (I) from a host organism. The host is transformed
CC   with the following: (i) a DNA sequence (Ia), flanked by 5' and 3'
CC   recombination DNA sequences (RS); and (ii) a sequence (II) that encodes a
CC   ligand-binding domain/recombinase fusion protein (Rec-LBD) that
CC   recognizes RS, under control of an inducible promoter (iP).
CC   Transformation is under conditions where iP is repressed (no Rec-LBD is
CC   produced), after which iP is induced to cause expression of Rec-LBD and
CC   a ligand is added to activate this fusion protein. The method is
CC   applicable to plants to provide temporally regulated expression of
CC   foreign proteins ('protein farming'), particularly for post-harvest
CC   recovery of the proteins. The proteins are particularly toxic or
CC   deleterious to the plant, e.g. antimicrobial or antibiotic peptides,
CC   RNAases, diphtheria toxin, cytosine deaminase, antibodies etc. The method
CC   may also be used to remove marker genes from transgenic plants and for
CC   production of poly(hydroxyalkanoates). The method includes two levels of
CC   repression (use of inducible promoter and ligand for activating the
CC   recombination reaction), ensuring secure regulation of recombinase
CC   activity as long as this is required (generally until plants are
CC   harvested). By selection of appropriate promoters, tissue selectivity may
CC   also be provided. This sequence represents a linker primer used in the
CC   construction of plasmid pRT100/scFv(ox)E which contains an endoplasmic
CC   reticulum localising scFv(ox) antibody fragment encoding a KDEL-ER
CC   targetting sequence.
XX
SQ   Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

  Query Match             87.7%;  Score 11.4;  DB 24;  Length 14;
  Best Local Similarity   92.3%;  Pred. No. 2.6e+03;
  Matches   12;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGGCATGGCATG 13
           ||| |||||||||
Db       2 ATGCCATGGCATG 14


RESULT 12
AAI69334/c
ID   AAI69334 standard; DNA; 14 BP.
XX
AC   AAI69334;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Plasmid pRT100/scFv(ox)E ER localising scFv(ox) Ab linker primer #1.
XX
KW   Endoplasmic reticulum; scFv; antibody; protein farming; transgenic plant;
KW   antimicrobial; antibiotic; RNAase; diphtheria toxin; cytosine deaminase;
KW   polyhydroxyalkanoate production; primer; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag=  a
FT                   /mod_base= "OTHER"
FT                   /note= "5'-end phosphorylated"
XX
PN   WO200188164-A1.
XX
PD   22-NOV-2001.
XX
PF   28-FEB-2001; 2001WO-DE00780.
XX
PR   19-MAY-2000; 2000DE-1024740.
XX
PA   (MPBC-) MPB COLOGNE GMBH.
XX
PI   Duering K;
XX
DR   WPI; 2002-055703/07.
XX
PT   Controlled elimination of DNA, useful for expressing toxic proteins in
PT   plants, comprises expressing recombinase-ligand binding domain fusion
PT   from an inducible promoter  -
XX
PS   Example 1; Page 18; 31pp; German.
XX
CC   This invention describes a novel method for the controlled elimination of
CC   a selected DNA sequence (I) from a host organism. The host is transformed
CC   with the following: (i) a DNA sequence (Ia), flanked by 5' and 3'
CC   recombination DNA sequences (RS); and (ii) a sequence (II) that encodes a
CC   ligand-binding domain/recombinase fusion protein (Rec-LBD) that
CC   recognizes RS, under control of an inducible promoter (iP).
CC   Transformation is under conditions where iP is repressed (no Rec-LBD is
CC   produced), after which iP is induced to cause expression of Rec-LBD and
CC   a ligand is added to activate this fusion protein. The method is
CC   applicable to plants to provide temporally regulated expression of
CC   foreign proteins ('protein farming'), particularly for post-harvest
CC   recovery of the proteins. The proteins are particularly toxic or
```

CC deleterious to the plant, e.g. antimicrobial or antibiotic peptides,
CC RNAases, diphtheria toxin, cytosine deaminase, antibodies etc. The method
CC may also be used to remove marker genes from transgenic plants and for
CC production of poly(hydroxyalkanoates). The method includes two levels of
CC repression (use of inducible promoter and ligand for activating the
CC recombination reaction), ensuring secure regulation of recombinase
CC activity as long as this is required (generally until plants are
CC harvested). By selection of appropriate promoters, tissue selectivity may
CC also be provided. This sequence represents a linker primer used in the
CC construction of plasmid pRT100/scFv(ox)E which contains an endoplasmic
CC reticulum localising scFv(ox) antibody fragment encoding a KDEL-ER
CC targetting sequence.
XX
SQ Sequence 14 BP; 3 A; 4 C; 4 G; 3 T; 0 other;

Query Match 87.7%; Score 11.4; DB 24; Length 14;
Best Local Similarity 92.3%; Pred. No. 2.6e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||| |||||||||
Db 13 ATGCCATGGCATG 1

RESULT 13
AAZ70542/c
ID AAZ70542 standard; DNA; 18 BP.
XX
AC AAZ70542;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human biallelic marker upstream amplification primer SEQ ID NO:4898.
XX
KW Human genome; biallelic marker; high density disequilibrium map;
KW genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW haplotyping; hybridisation; identification; characterisation;
KW amplification; single nucleotide polymorphism; SNP; PCR primer;
KW diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9954500-A2.
XX
PD 28-OCT-1999.
XX
PF 21-APR-1999; 99WO-IB00822.
XX
PR 21-APR-1998; 98US-0082614.
PR 23-NOV-1998; 98US-0109732.
XX
PA (GEST) GENSET.
XX
PI Cohen D, Blumenfeld M, Chumakov I;
XX
DR WPI; 2000-013267/01.
XX
PT Novel biallelic markers used to construct a high density disequilibrium
PT map of the human genome -
XX
PS Claim 8; Page 1275; 2745pp; English.
XX
CC AAZ65654 to AAZ69578 represent human biallelic markers from the present
CC invention, which contain a polymorphic base at position 24 of their
CC nucleotide sequences. AAZ69579 to AAZ77440 represent amplification
CC primers for the biallelic markers. The biallelic markers of the
CC invention have a variety of uses: they can be used for high density
CC mapping of the human genome, and in complex association studies and
CC haplotyping studies which are useful in determining the genetic basis
CC for disease states. Compositions and methods of the invention can also
CC be useful for the identification of the targets for the development of
CC pharmaceutical agents and diagnostic methods, as well as the
CC characterisation of the differential efficacious responses to and side
CC effects from pharmaceutical agents acting on a disease as well as other
CC treatment.
CC N.B. The SEQ ID NOs 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
CC and 3367, are not actually given a sequence in the Sequence Listing
CC from the present invention.
XX
SQ Sequence 18 BP; 4 A; 5 C; 4 G; 5 T; 0 other;

Query Match 87.7%; Score 11.4; DB 21; Length 18;
Best Local Similarity 92.3%; Pred. No. 2.7e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||||||| |||||
Db 13 ATGGCATTGCATG 1

RESULT 14
AAT39773
ID AAT39773 standard; DNA; 20 BP.
XX
AC AAT39773;
XX
DT 30-APR-1997 (first entry)
XX
DE Target sequence for p53 peptide binding.
XX
KW Human; p53; cell proliferation; cell death; regulator; tumour; psoriasis;
KW negative regulatory region; DNA damaging agent; transplant rejection;
KW abnormal cell proliferation; atherosclerosis; cancer; autoimmune disease;
KW arterial restenosis; immune response; apoptosis; inducer; therapy;
KW proliferating lymphocytes; ds.
XX
OS Synthetic.
XX
PN WO9625434-A1.
XX
PD 22-AUG-1996.
XX
PF 16-FEB-1996; 96WO-US01535.
XX
PR 16-FEB-1995; 95US-0392542.
XX
PA (FARB) BAYER CORP.
PA (WIST-) WISTAR INST.
XX
PI Halazonetis T, Hartwig W;
XX
DR WPI; 1996-393345/39.
XX
PT New human p53-isomorphic peptide(s) and peptido:mimetic cpds. - used
PT for activating p53 function, e.g. for treating tumours, cancers,
PT psoriasis, etc
XX
PS Claim 21; Page 47; 55pp; English.
XX
CC This sequence represents a target sequence used in a DNA binding assay
CC to detect p53 mutants whose DNA binding ability is activated by a peptide
CC of the invention (see AAW05350-W05364). The peptides of the invention
CC consist of at least four sequential amino acids from a negative
CC regulatory region which maps to residues 361-383 of p53 (see AAW05344 for
CC wild type p53 sequence). The p53 protein functions to regulate cell
CC proliferation and cell death, and is mutated in more than half of all
CC human tumours. The peptide sequences preferably contain four amino acids
CC from a non-human p53 sequence, contain D-form amino acids, and can also
CC be cyclic peptides. The sequences retain the structural characteristics
CC of the original peptides, but the modifications render them less
CC susceptible to cleavage by proteases and exopeptidases. As the peptides
CC activate p53 DNA binding, they can be used to identify p53 mutants. The
CC peptides can also be used for treating a patient with a tumour expressing
CC a p53 mutant whose ability to bind DNA may be activated by one of the
CC peptides. They can also be used for treating conditions such as exposure
CC to DNA damaging agents, abnormal cell proliferation characteristic of

CC psoriasis, atherosclerosis, cancer, arterial restenosis, autoimmune
CC diseases and undesirable immune responses accompanying rejection of a
CC transplant. The peptides can also induce apoptosis of specific cells,
CC such as proliferating lymphocytes.
XX
SQ Sequence 20 BP; 4 A; 4 C; 6 G; 6 T; 0 other;

Query Match 87.7%; Score 11.4; DB 17; Length 20;
Best Local Similarity 92.3%; Pred. No. 2.7e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||| |||||||||
Db 5 ATGTCATGGCATG 17

RESULT 15
AAT07059/c
ID AAT07059 standard; DNA; 20 BP.
XX
AC AAT07059;
XX
[redacted] 05-JUL-1996 (first entry)

DE Primer 17A-215F for n-(ABCDE) hepatitis virus.
XX
KW Oligonucleotide 5' primer 17A-215F; polymerase chain reaction;
KW PCR; non-A, non-B, non-C, non-D, non-E hepatitis virus;
KW n-(ABCDE); immunogen; antibody; vaccine; phage library; ss.
XX
OS Synthetic.
XX
PN WO9532290-A2.
XX
PD 30-NOV-1995.
XX
PF 17-MAY-1995; 95WO-US05980.
XX
PR 20-MAY-1994; 94US-0246986.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Kim JP, Purcell RH;
XX
DR WPI; 1996-020585/02.
XX
PT New non-A, -B, -C, -D and -E (n-(ABCDE)) hepatitis DNA libraries -
PT used to develop prods. for the detection, diagnosis, prevention and
PT treatment of n-(ABCDE) hepatitis.
XX
[redacted] Example 10; Page 126; 165pp; English.
XX
CC After immunoscreening of the JFA library (ATCC 75271) in phage
CC lambda gt11 for the isolation of clones encoding immunogenic
CC polypeptides associated with non-A, non-B, non-C, non-D, non-E
CC (n-(ABCDE)) hepatitis virus infection, overlapping clones to clone
CC 17A (see AAT07037) were obtained. One such clone, WT54 (AAT07253),
CC contained a 210 bp overlap with 17A which extends 119 bp from the
CC 3' end of 17A. This 5' primer is used with 3' WT54-684R (AAT07058)
CC to determine that the 17A-WT54 linked sequence is present in JFA
CC DNA and is not an artifact. n-(ABCDE) hepatitis polypeptides can
CC be used for the production or detection of antibodies, and in
CC vaccines. The antibodies can be used for detection, diagnosis and
CC in passive immunotherapy. The DNA can also be used in detection
CC and diagnosis, and as hybridisation probes for identification of
CC further n-(ABCDE) hepatitis coding sequences. Culture systems
CC producing the n-(ABCDE) polypeptides can be used in screening
CC studies.
XX
SQ Sequence 20 BP; 5 A; 7 C; 4 G; 4 T; 0 other;

Query Match 87.7%; Score 11.4; DB 17; Length 20;
Best Local Similarity 92.3%; Pred. No. 2.7e+03;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||||||||||| ||
Db 14 ATGGCATGGCTTG 2

Search completed: March 29, 2003, 23:57:01
Job time : 3577 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 22:48:37 ; Search time 29060 Seconds
(without alignments)
7.245 Million cell updates/sec

Title: US-09-897-776A-18
Perfect score: 13
Sequence: 1 atggcatggcatg 13

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 102592

Minimum DB seq length: 13
Maximum DB seq length: 50

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	11	84.6	28	17	AZ412849	AZ412849 1M0186K06
	2	11	84.6	35	10	BE618566	BE618566 601462356
c	3	10.4	80.0	27	17	AZ448611	AZ448611 1M0246H21
c	4	10.4	80.0	27	17	TA385H06Q	AL498874 T. brucei
	5	10.4	80.0	30	17	AZ841194	AZ841194 2M0139L03
	6	10.4	80.0	31	17	AZ764843	AZ764843 1M0561N21
	7	10.4	80.0	46	17	AQ254727	AQ254727 EP(3)3520
	8	10.4	80.0	49	9	AA954745	AA954745 on56a04.s
c	9	10	76.9	22	9	AI802216	AI802216 tj36b07.x
	10	10	76.9	23	17	AZ588147	AZ588147 1M0396G12
c	11	10	76.9	34	9	AA630482	AA630482 ab98f09.s
c	12	10	76.9	44	17	B07611	B07611 CDC10d4 Cri
	13	10	76.9	46	9	AA444354	AA444354 vd92c08.r
c	14	9.8	75.4	23	17	AZ799538	AZ799538 2M0057N05
c	15	9.8	75.4	28	17	AZ361511	AZ361511 1M0106B17
	16	9.8	75.4	32	17	AZ657897	AZ657897 1M0534F19
	17	9.8	75.4	36	17	AZ797263	AZ797263 2M0053G23
c	18	9.8	75.4	36	17	TA343G09P	AL493727 T. brucei
c	19	9.8	75.4	37	14	H40236	H40236 yn85h10.r1
	20	9.8	75.4	37	17	AZ366190	AZ366190 1M0115M19
	21	9.8	75.4	38	17	AZ850114	AZ850114 2M0151P18
c	22	9.8	75.4	40	9	AI040658	AI040658 ox23g04.s
	23	9.8	75.4	43	9	AA663711	AA663711 ae71a03.s
	24	9.8	75.4	47	17	AL760555	AL760555 Arabidops
c	25	9.8	75.4	48	13	BJ050453	BJ050453 BJ050453
	26	9.8	75.4	49	10	AV964401	AV964401 AV964401
c	27	9.8	75.4	50	9	AU103603	AU103603 AU103603
	28	9.4	72.3	20	17	AZ442667	AZ442667 1M0237C01
	29	9.4	72.3	27	17	AZ776662	AZ776662 2M0010C06
c	30	9.4	72.3	27	17	AZ966603	AZ966603 2M0237J12
c	31	9.4	72.3	29	17	AZ603273	AZ603273 1M0422B06
	32	9.4	72.3	31	17	BH862004	BH862004 SALK_0885
	33	9.4	72.3	40	13	BI770067	BI770067 603053595
c	34	9.4	72.3	42	17	BH855809	BH855809 SALK_0845
c	35	9.4	72.3	44	17	AZ615083	AZ615083 1M0444B17
c	36	9.4	72.3	44	17	BH855815	BH855815 SALK_0845
c	37	9.4	72.3	46	9	AA411525	AA411525 zv27c09.s
	38	9.4	72.3	46	9	AA531890	AA531890 TgESTzz32
c	39	9.4	72.3	49	9	AA927087	AA927087 oi58a06.s
c	40	9.4	72.3	49	9	AI003826	AI003826 ot43f04.s
c	41	9.4	72.3	49	9	AA122130	AA122130 zn82f09.r
c	42	9.4	72.3	50	9	AU102369	AU102369 AU102369
c	43	9.4	72.3	50	9	AU102370	AU102370 AU102370
	44	9.4	72.3	50	9	AU104846	AU104846 AU104846
c	45	9.4	72.3	50	9	AU105888	AU105888 AU105888

ALIGNMENTS

RESULT 1
AZ412849
LOCUS AZ412849 28 bp DNA linear GSS 03-OCT-2000
DEFINITION 1M0186K06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0186K06 R, DNA sequence.
ACCESSION AZ412849
VERSION AZ412849.1 GI:10536862
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 28)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

```
            Plate: 0186  row: K  column: 06
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES              Location/Qualifiers
     source           1. .28
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0186K06"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT        8 a      6 c      7 g      7 t
ORIGIN

  Query Match             84.6%;  Score 11;  DB 17;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 4e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGGCATGGCA 11
          ||||||||||
Db     17 ATGGCATGGCA 27


RESULT 2
BE618566
LOCUS       BE618566                 35 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601462356T1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3865957 3',
            mRNA sequence.
ACCESSION   BE618566
VERSION     BE618566.1  GI:9889504
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9609  row: o  column: 14
            High quality sequence stop: 35.
FEATURES              Location/Qualifiers
     source           1. .35
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3865957"
                      /clone_lib="NIH_MGC_67"
                      /tissue_type="retinoblastoma"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.75 kb.   Library constructed by Life
                      Technologies."
BASE COUNT        6 a      8 c      6 g     15 t
ORIGIN

  Query Match             84.6%;  Score 11;  DB 10;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TGGCATGGCAT 12
          ||||||||||
Db      9 TGGCATGGCAT 19


RESULT 3
AZ448611/c
LOCUS       AZ448611                 27 bp    DNA     linear   GSS 04-OCT-2000
DEFINITION  1M0246H21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0246H21 F, DNA sequence.
ACCESSION   AZ448611
VERSION     AZ448611.1  GI:10601577
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0246  row: H  column: 21
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 27.
FEATURES              Location/Qualifiers
     source           1. .27
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0246H21"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
```

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 8 a 12 c 5 g 2 t
ORIGIN

Query Match 80.0%; Score 10.4; DB 17; Length 27;
Best Local Similarity 91.7%; Pred. No. 7.8e+04;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TGGCATGGCATG 13
|| |||||||||
Db 15 TGTCATGGCATG 4

RESULT 4
TA385H06Q/c
LOCUS TA385H06Q 27 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 385h06, reverse sequence, genomic survey sequence.
ACCESSION AL498874
VERSION AL498874.1 GI:11874596
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 27)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh1@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (
4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES Location/Qualifiers
source 1. .27
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="385h06"
BASE COUNT 12 a 7 c 3 g 5 t
ORIGIN

Query Match 80.0%; Score 10.4; DB 17; Length 27;
Best Local Similarity 91.7%; Pred. No. 7.8e+04;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCAT 12
|||||||||| ||
Db 15 ATGGCATGGAAT 4

RESULT 5
AZ841194
LOCUS AZ841194 30 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0139L03F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0139L03 F, DNA sequence.
ACCESSION AZ841194
VERSION AZ841194.1 GI:13011102
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 30)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0139 row: L column: 03
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 30.
FEATURES Location/Qualifiers
source 1. .30
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0139L03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 6 a 2 c 12 g 10 t
ORIGIN

Query Match 80.0%; Score 10.4; DB 17; Length 30;
Best Local Similarity 91.7%; Pred. No. 8.2e+04;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TGGCATGGCATG 13
|||||||| |||
Db 11 TGGCATGGGATG 22

```
RESULT 6
AZ764843
LOCUS       AZ764843                 31 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  1M0561N21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0561N21 F, DNA sequence.
ACCESSION   AZ764843
VERSION     AZ764843.1  GI:12880199
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 31)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0561  row: N  column: 21
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 31.
FEATURES             Location/Qualifiers
     source          1. .31
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0561N21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       12 a      5 c      5 g      9 t
ORIGIN

  Query Match             80.0%;  Score 10.4;  DB 17;  Length 31;
  Best Local Similarity   91.7%;  Pred. No. 8.3e+04;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TGGCATGGCATG 13
           || |||||||||
Db      10 TGACATGGCATG 21
```

```
RESULT 7
AQ254727
LOCUS       AQ254727                 46 bp    DNA     linear   GSS 23-AUG-2000
DEFINITION  EP(3)3520 Drosophila melanogaster EP line Drosophila melanogaster
            genomic Sequence recovered from Both 5' and 3' ends of P element,
            DNA sequence.
ACCESSION   AQ254727
VERSION     AQ254727.1  GI:3738590
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Liao,G.-C., Rehm,E.J. and Rubin,G.M.
  TITLE     Insertion site preferences of the P transposable element in
            Drosophila melanogaster
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3347-3351 (2000)
  MEDLINE   20202638
COMMENT     Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu
            Sequence recovery method was inverse PCR.

            Sequence orientation is forward strand relative to 5' end of P
            element

            The P element insertion position is base 17 in the 46 bases. This
            insertion position refers to the first base of the 8 base target
            recognition sequence.
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="Drosophila melanogaster EP line"
                     /note="Inverse PCR was performed on Drosophila
                     melanogaster strains each of which contains a single EP
                     transposable element insertion. (The generation of these
                     insertion strains is described in Rorth P, Szabo K, Bailey
                     A, Laverty T, Rehm J, Rubin GM, Weigmann K, Milan M, Benes
                     V, Ansorge W, Cohen SM. 1998. Systematic gain-of-function
                     genetics in Drosophila. Development 6:1049-1057.) The
                     resultant fragment for each strain was directly sequenced
                     to determine the genomic sequence at the site of
                     insertion. Details of the protocols used can be found at
                     http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT       10 a     13 c     13 g     10 t
ORIGIN

  Query Match             80.0%;  Score 10.4;  DB 17;  Length 46;
  Best Local Similarity   91.7%;  Pred. No. 9.8e+04;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TGGCATGGCATG 13
           |||||||||  ||
Db      10 TGGCATGGCCTG 21
```

```
RESULT 8
AA954745
LOCUS       AA954745                 49 bp    mRNA    linear   EST 23-JUN-1998
DEFINITION  on56a04.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1560654 3' similar to gb:M21248 GLUCOSE-6-PHOSPHATE
            1-DEHYDROGENASE (HUMAN);, mRNA sequence.
ACCESSION   AA954745
VERSION     AA954745.1  GI:3118440
KEYWORDS    EST.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1780  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1560654"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT        7 a      4 c     21 g     17 t
ORIGIN

  Query Match             80.0%;  Score 10.4;  DB 9;  Length 49;
  Best Local Similarity   91.7%;  Pred. No. 1e+05;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGGCATGGCAT 12
           |||||| |||||
Db      22 ATGGCAGGGCAT 33


[redacted]LT 9
AI802216/c
LOCUS       AI802216                 22 bp    mRNA    linear   EST 13-DEC-1999
DEFINITION  tj36b07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2143573 3'
            similar to SW:RL3_BOVIN P39872 60S RIBOSOMAL PROTEIN L3. ;contains
            PTR5.b2 PTR5 repetitive element ;, mRNA sequence.
ACCESSION   AI802216
VERSION     AI802216.1  GI:5367688
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1000  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.

FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2143573"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT        5 a      9 c      6 g      2 t
ORIGIN

  Query Match             76.9%;  Score 10;  DB 9;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GGCATGGCAT 12
           ||||||||||
Db      20 GGCATGGCAT 11


RESULT 10
AZ588147
LOCUS       AZ588147                 23 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0396G12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0396G12 F, DNA sequence.
ACCESSION   AZ588147
VERSION     AZ588147.1  GI:11710253
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0396  row: G  column: 12
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 23.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0396G12"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
```

polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT        8 a      7 c      4 g      4 t
ORIGIN

  Query Match             76.9%;  Score 10;  DB 17;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GCATGGCATG 13
           ||||||||||
Db      13 GCATGGCATG 22


RESULT 11
AA630482/c
LOCUS       AA630482                 34 bp    mRNA    linear   EST 06-MAR-1998
DEFINITION  ab98f09.s1 Stratagene lung (#937210) Homo sapiens cDNA clone
            IMAGE:855017 3' similar to SW:RL3_HUMAN P39023 60S RIBOSOMAL
            PROTEIN L3. ;, mRNA sequence.
ACCESSION   AA630482
VERSION     AA630482.1  GI:2553093
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 34)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Insert Length: 725   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .34
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:855017"
                     /clone_lib="Stratagene lung (#937210)"
                     /sex="male"
                     /dev_stage="72 years"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. normal lung.  Average insert size: 1.0 kb; Uni-ZAP XR
                     Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       10 a     15 c      6 g      3 t
ORIGIN

  Query Match             76.9%;  Score 10;  DB 9;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GGCATGGCAT 12
           ||||||||||
Db      20 GGCATGGCAT 11


RESULT 12
B07611/c
LOCUS       B07611                   44 bp    DNA     linear   GSS 17-JAN-1998
DEFINITION  CDC10d4 Cri du chat, exon trapped products Homo sapiens genomic
            clone CDC10d4, DNA sequence.
ACCESSION   B07611
VERSION     B07611.1  GI:2781450
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 44)
  AUTHORS   Church,D.M., Yang,J., Shiang,R., Wasmuth and J.J.
  TITLE     A High Resolution Physical and Transcription map for the Cri du
            chat region of human chromosome 5p
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Rita Shiang
            Shiang lab
            University of California- Irvine
            240D Med. Sci. I, UCI-COM, Irvine, CA 92697-1700, USA
            Tel: (714)824-6792
            Fax: (714)824-3403
            Email: rshiang@chrom5.hsis.uci.edu
            Insert Length: 44   Std Error: 0.00
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .44
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CDC10d4"
                     /clone_lib="Cri du chat, exon trapped products"
                     /note="Exon trapped products from the CDC critical region
                     associated with mental retardation and facial
                     dysmorphism."
BASE COUNT       10 a     13 c     14 g      7 t
ORIGIN

  Query Match             76.9%;  Score 10;  DB 17;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GGCATGGCAT 12
           ||||||||||
Db      22 GGCATGGCAT 13


RESULT 13
AA444354
LOCUS       AA444354                 46 bp    mRNA    linear   EST 03-JUN-1997
DEFINITION  vd92c08.r1 Soares mouse NbMH Mus musculus cDNA clone IMAGE:808046
            5' similar to SW:ATPJ_MOUSE Q06185 ATP SYNTHASE E CHAIN,
            MITOCHONDRIAL ;, mRNA sequence.
ACCESSION   AA444354
VERSION     AA444354.1  GI:2157019
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
```

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:484390
Putative full length read
vector to vector length is 246
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 32.
FEATURES Location/Qualifiers
source 1. .46
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:808046"
/clone_lib="Soares mouse NbMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAAAGTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 8 a 16 c 13 g 9 t
ORIGIN

[redacted]ery Match 76.9%; Score 10; DB 9; Length 46;
[redacted]st Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GGCATGGCAT 12
||||||||||
Db 23 GGCATGGCAT 32

RESULT 14
AZ799538/c
LOCUS AZ799538 23 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0057N05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0057N05 F, DNA sequence.
ACCESSION AZ799538
VERSION AZ799538.1 GI:12950757
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 23)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0057 row: N column: 05
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 23.
FEATURES Location/Qualifiers
source 1. .23
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0057N05"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 6 a 5 c 4 g 8 t
ORIGIN

Query Match 75.4%; Score 9.8; DB 17; Length 23;
Best Local Similarity 84.6%; Pred. No. 1.5e+05;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCATGGCATG 13
||||||| | |||
Db 18 ATGGCATCGTATG 6

RESULT 15
AZ361511/c
LOCUS AZ361511 28 bp DNA linear GSS 02-OCT-2000
DEFINITION 1M0106B17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0106B17 F, DNA sequence.
ACCESSION AZ361511
VERSION AZ361511.1 GI:10475211
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 28)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0106 row: B column: 17
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 28.

FEATURES Location/Qualifiers
source 1. .28
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0106B17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 13 a 5 c 3 g 7 t
ORIGIN

Query Match 75.4%; Score 9.8; DB 17; Length 28;
Best Local Similarity 84.6%; Pred. No. 1.6e+05;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 ATGGCATGGCATG 13
           ||||||| ||||
Db      15 ATGGCATTCCATG 3
```

Search completed: March 30, 2003, 08:21:31
Job time : 29063 secs